



EDA CASE STUDY

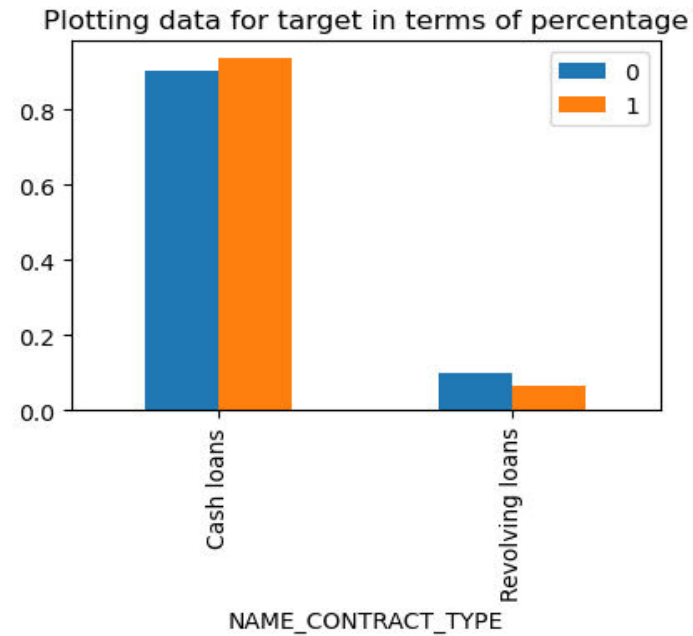
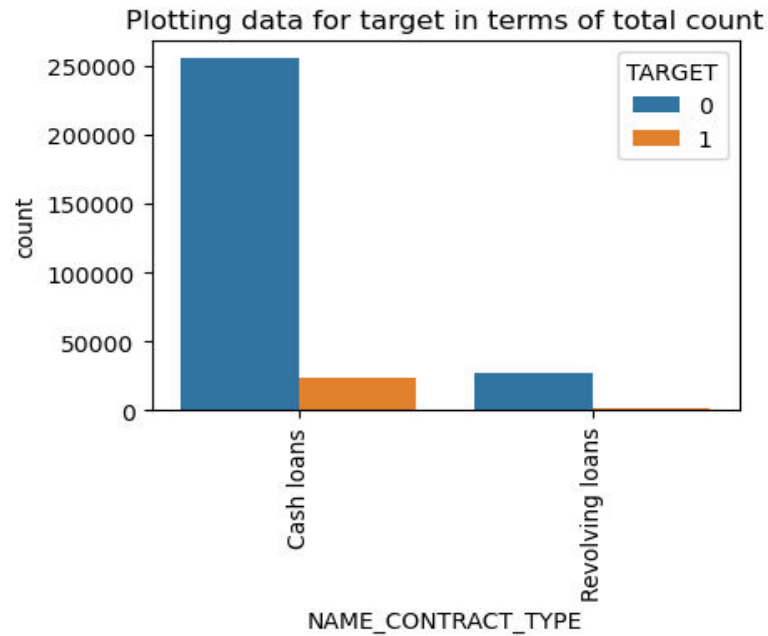
SINIYYA YOUSUF

Business Objectives

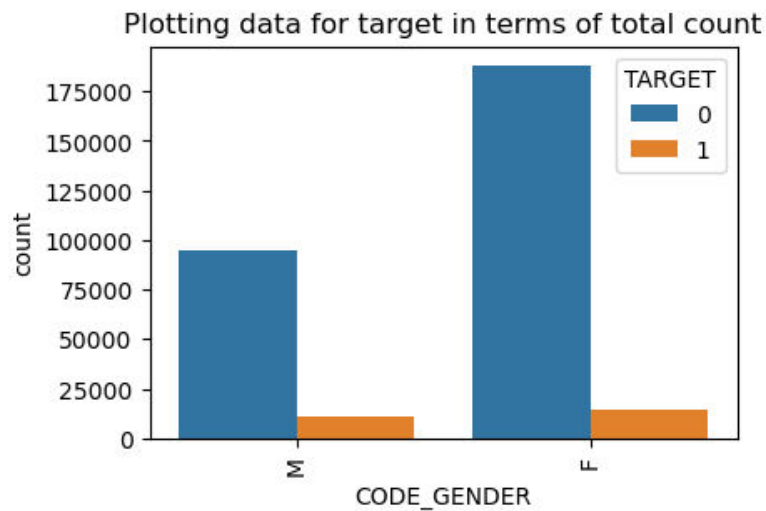
This case study aims to identify patterns which indicate if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Univariate Analysis on Categorical Variables (NAME_CONTRACT_TYPE, CODE_GENDER)

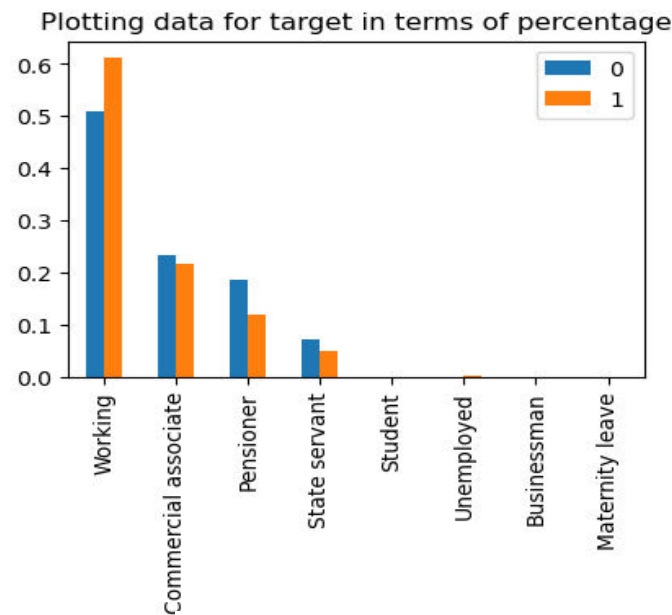
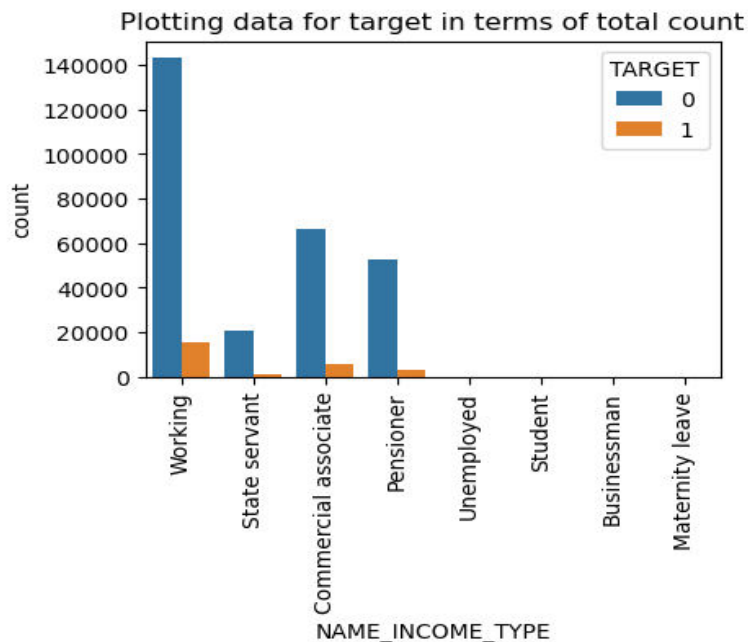


- **NAME_CONTRACT_TYPE**- Cash Loans are provided more than revolving loans ,about 90% are cash loans in both TARGET 1 and 0.



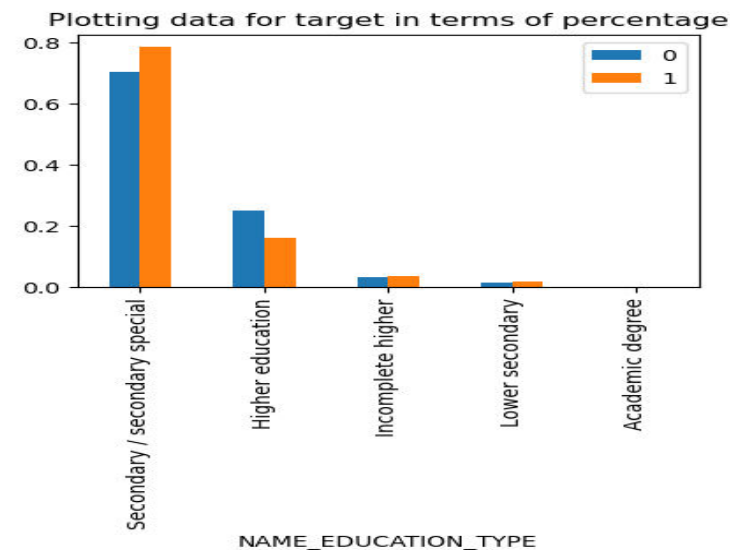
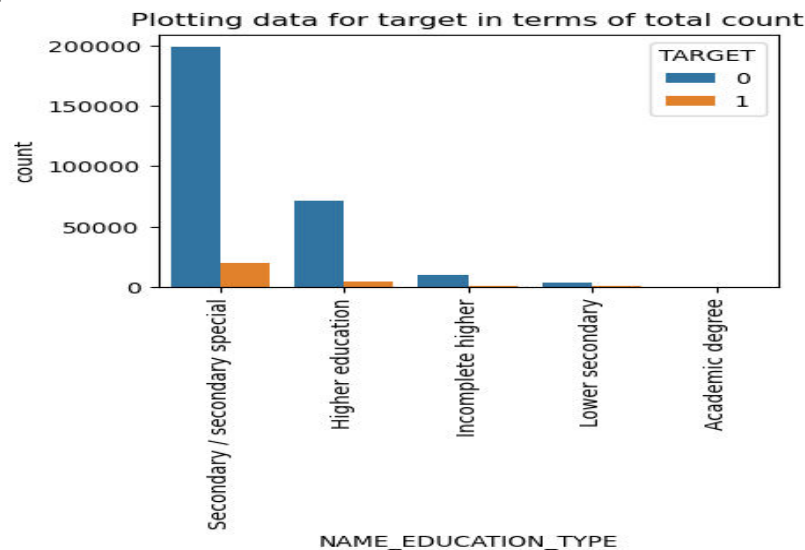
- **CODE_GENDER**- Even though Female applicants are more in number, the male applicants seems to be more likely to default.

Univariate Analysis on Categorical Variables (NAME_INCOME_TYPE & NAME_EDUCATION_TYPE)

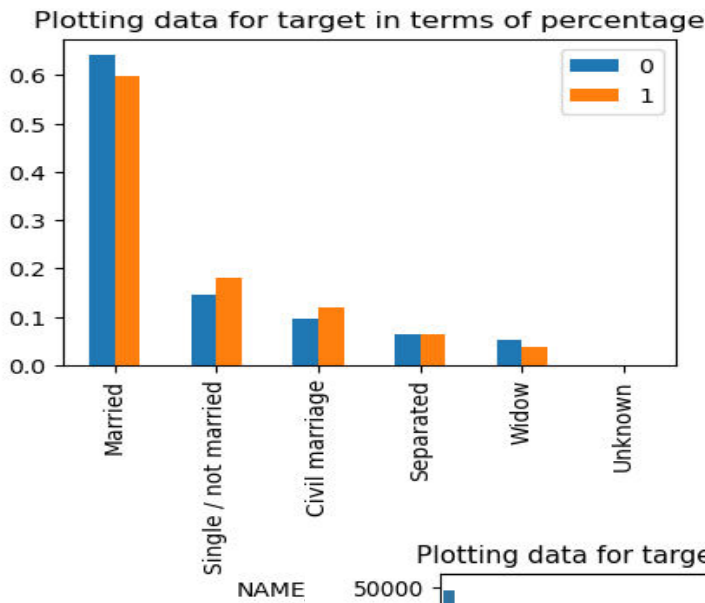
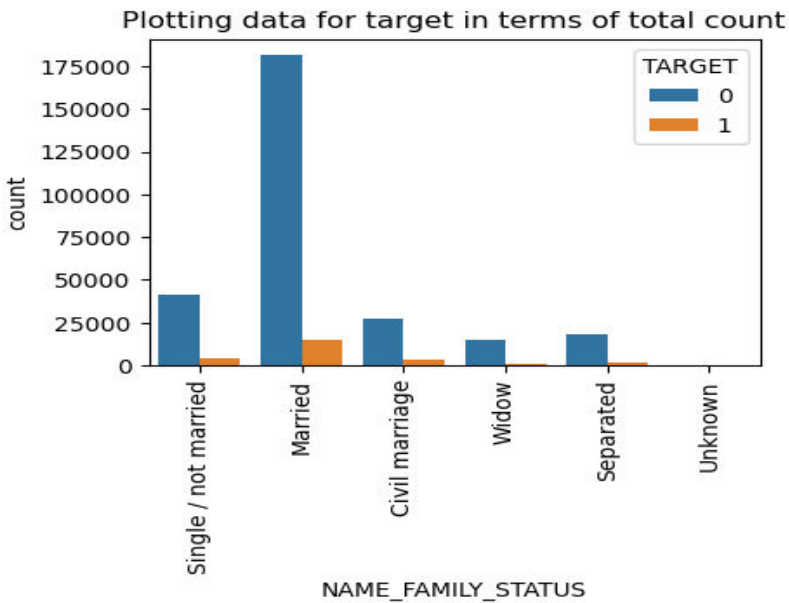


- NAME_INCOME_TYPE-Working class are the people who applies loan the most and there are significant percentage of defaulters present. Pensioners seem to be less defaulting considering their count in total application.

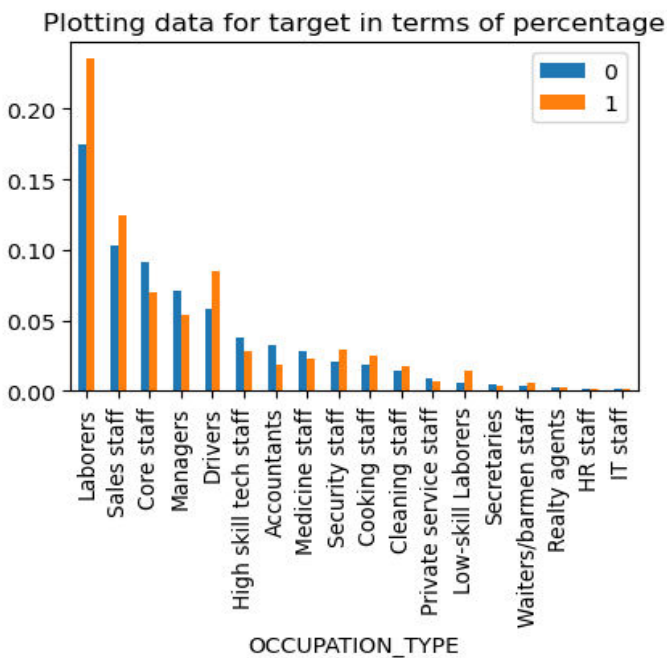
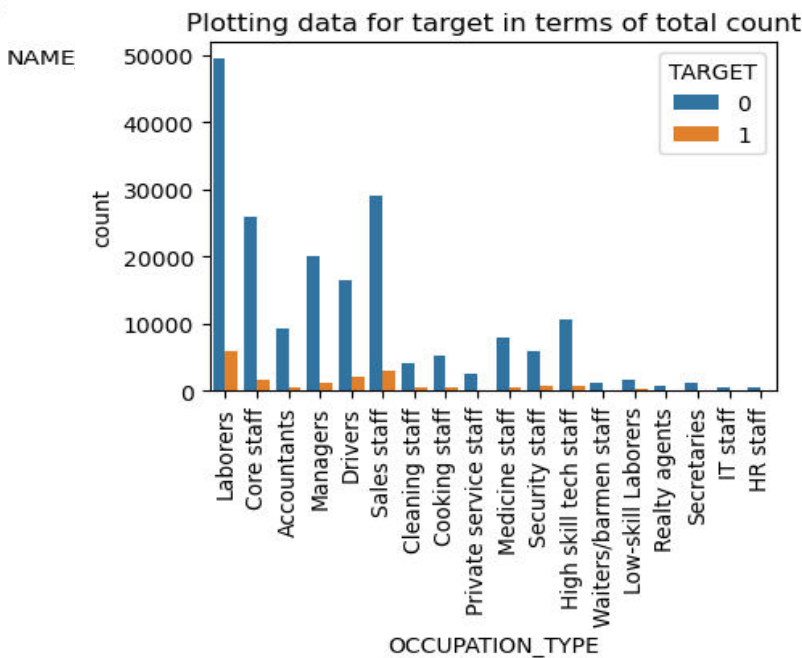
- NAME_EDUCATION_TYPE-Almost 70% of loan applicants have secondary education in TARGET 1 and near to 80% for TARGET 0.



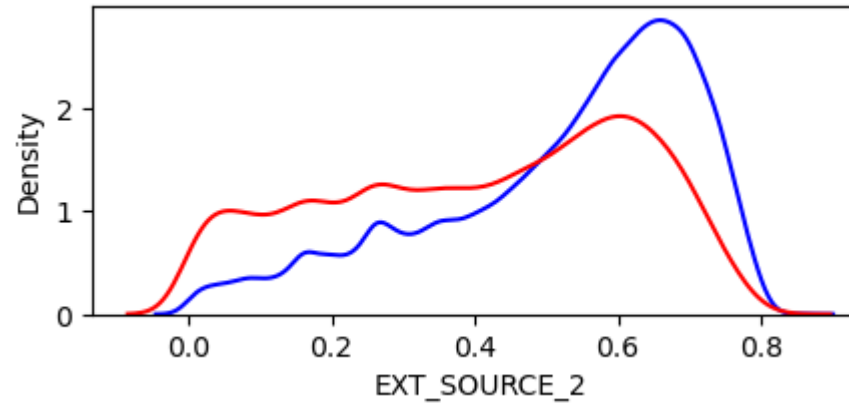
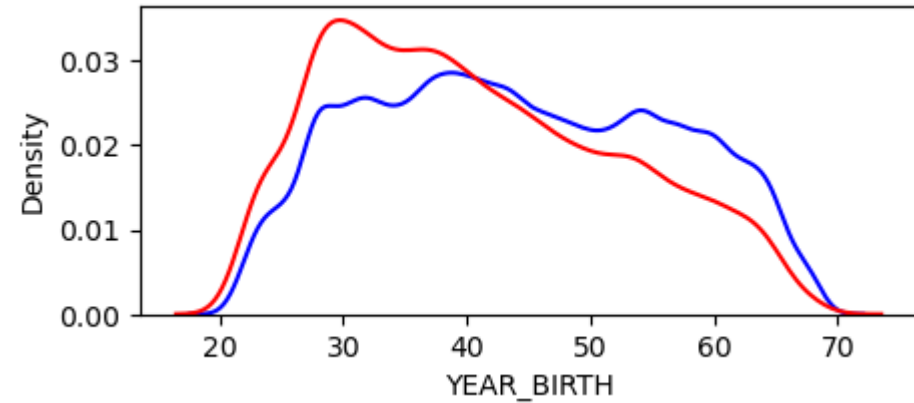
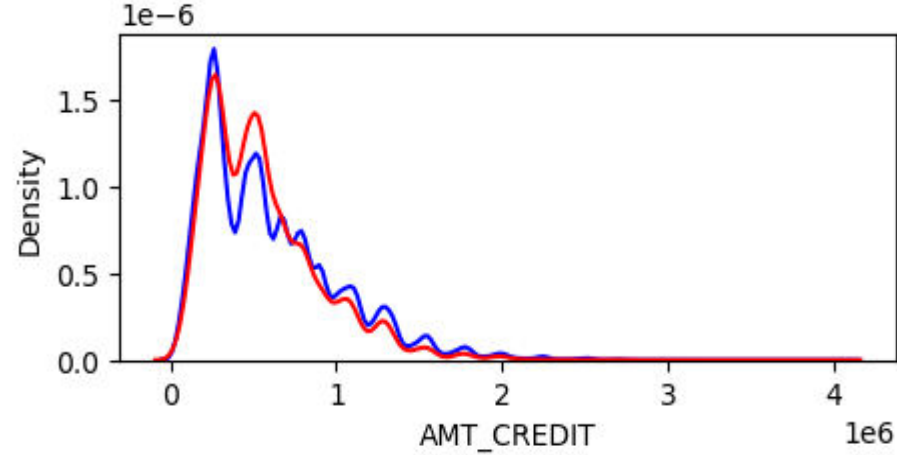
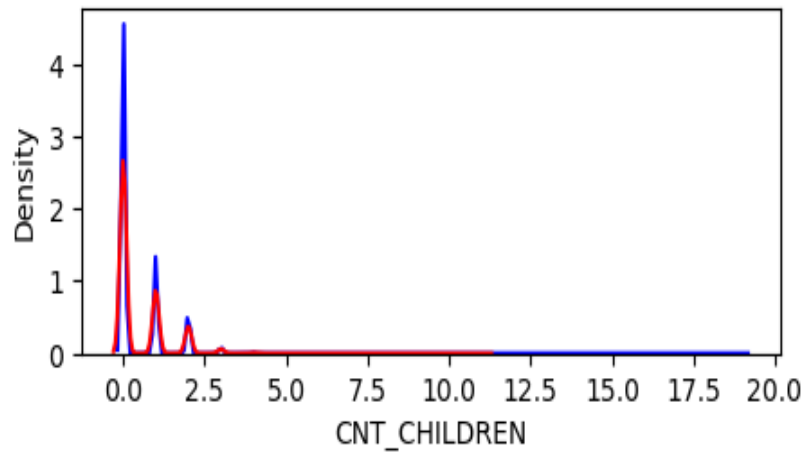
Univariate Analysis on Categorical Variables (NAME_FAMILY_STATUS ,OCCUPATION_TYPE)



- **NAME_FAMILY_STATUS**- Married people applies loan the most and there are significant number of defaulters too
- **OCCUPATION_TYPE**-Labourers and Sales staff are having more difficulty in repaying the loan.

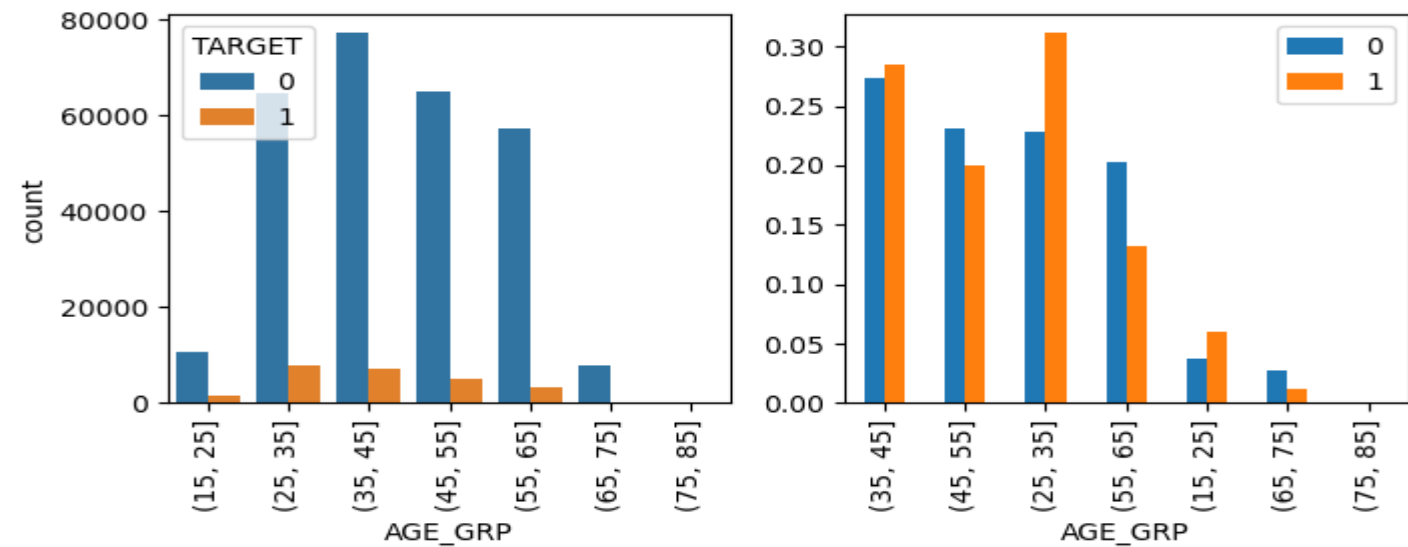


Univariate Analysis on Numerical variables

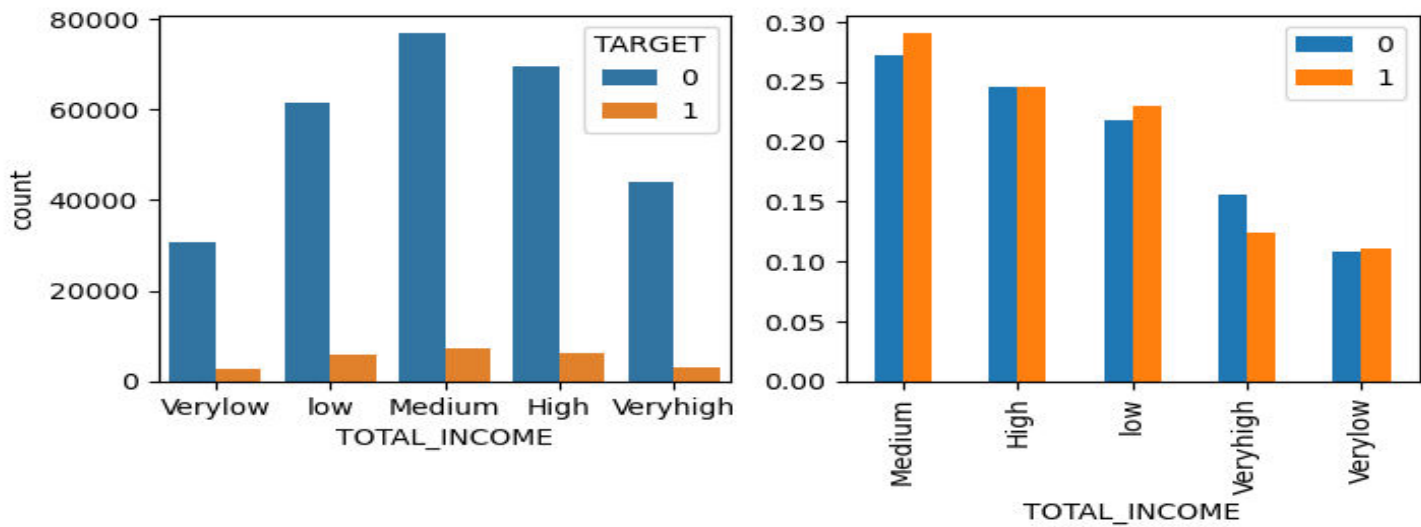


- **CNT_CHILDREN**-Both defaulters and non defaulters are more in number with 0 children but non defaulters seems to have more in number with 0 kids
- **AMT_CREDIT**-Amount credited to TARGET 1 seems to be slightly less compared to TARGET 0 which is good.
- **YEAR_BIRTH**- Defaulters seems to be around 30 years as density is more for TARGET 1.
- **EXT_SOURCE_2**- Normalized score from external data source is less for TARGET 0 which is expected.

Analysing AGE_GRP and TOTAL_INCOME for TARGET 1 and 0.

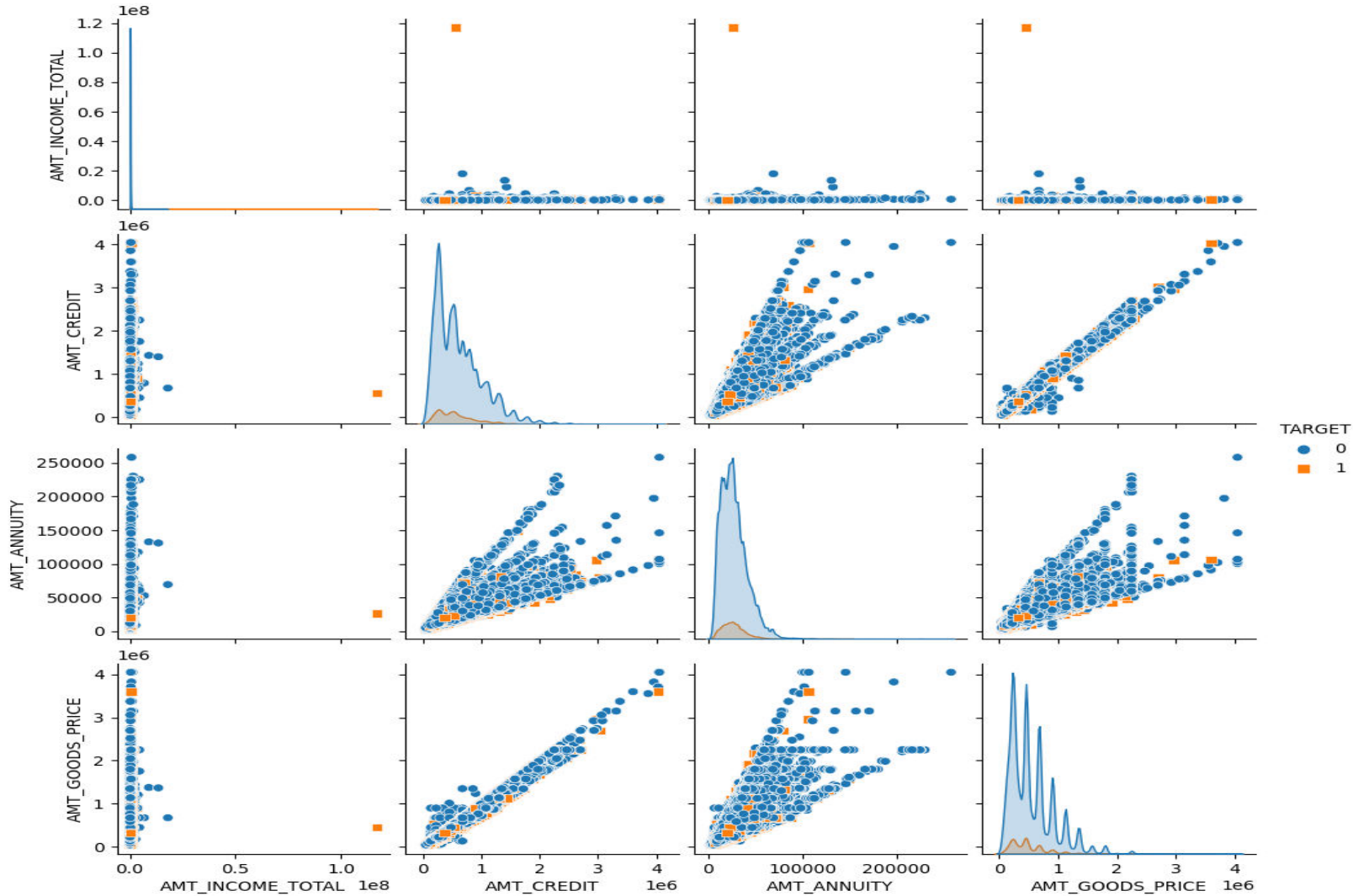


Defaulters seems to be more within age group 25-35 and 35-45.



Clients with Medium salary seems to be defaulting more than other income groups.

Bivariate Analysis of Numerical variables



- **AMT_CREDIT** and **AMT_GOODS_PRICE** have good correlation.
- **AMT_ANNUITY** also increases with **AMT_CREDIT**.
- **AMT_GOODS_PRICE** and **AMT_ANNUITY** also have positive correlation.

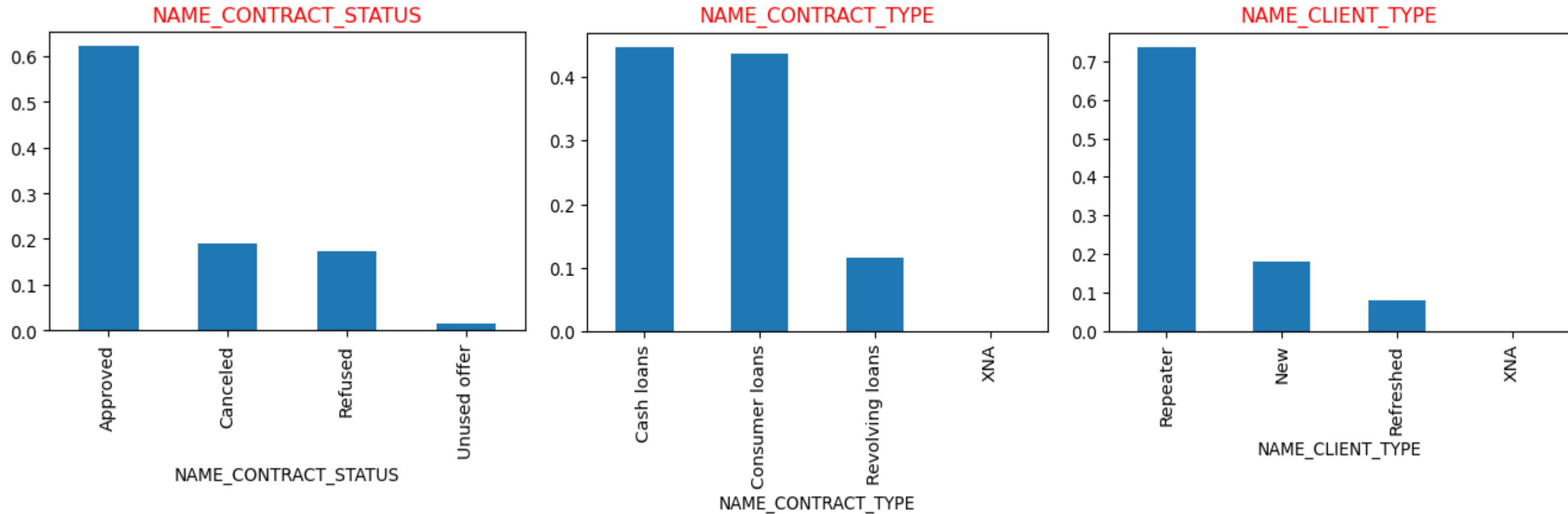
Correlation for Defaulters(TARGET = 1)

- OBS_30_CNT_SOCIAL_CIRCLE OBS_60_CNT_SOCIAL_CIRCLE 0.998269
- AMT_GOODS_PRICE AMT_CREDIT 0.983103
- REGION_RATING_CLIENT_W_CITY REGION_RATING_CLIENT 0.956637
- CNT_FAM_MEMBERS CNT_CHILDREN 0.885484
- DEF_60_CNT_SOCIAL_CIRCLE DEF_30_CNT_SOCIAL_CIRCLE 0.868994
- REG_REGION_NOT_WORK_REGION LIVE_REGION_NOT_WORK_REGION 0.847885
- REG_CITY_NOT_WORK_CITY LIVE_CITY_NOT_WORK_CITY 0.778540
- AMT_GOODS_PRICE AMT_ANNUITY 0.752699
- AMT_CREDIT AMT_ANNUITY 0.752195
- YEARS_EMPLOYED YEAR_BIRTH 0.582185

Correlation for Non Defaulters(TARGET = 0)

- OBS_60_CNT_SOCIAL_CIRCLE OBS_30_CNT_SOCIAL_CIRCLE 0.998508
- AMT_GOODS_PRICE AMT_CREDIT 0.987250
- REGION_RATING_CLIENT REGION_RATING_CLIENT_W_CITY 0.950149
- CNT_FAM_MEMBERS CNT_CHILDREN 0.878571
- REG_REGION_NOT_WORK_REGION LIVE_REGION_NOT_WORK_REGION 0.861861
- DEF_30_CNT_SOCIAL_CIRCLE DEF_60_CNT_SOCIAL_CIRCLE 0.859332
- REG_CITY_NOT_WORK_CITY LIVE_CITY_NOT_WORK_CITY 0.830381
- AMT_GOODS_PRICE AMT_ANNUITY 0.776686
- AMT_CREDIT AMT_ANNUITY 0.771309
- YEARS_EMPLOYED YEAR_BIRTH 0.626114

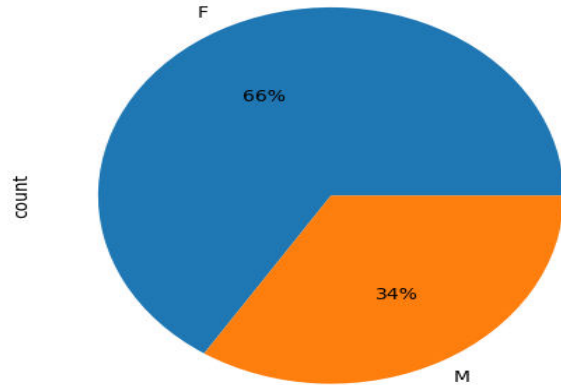
Analysis on NAME_CONTRACT_STATUS','NAME_CONTRACT_TYPE','NAME_CLIENT_TYPE'



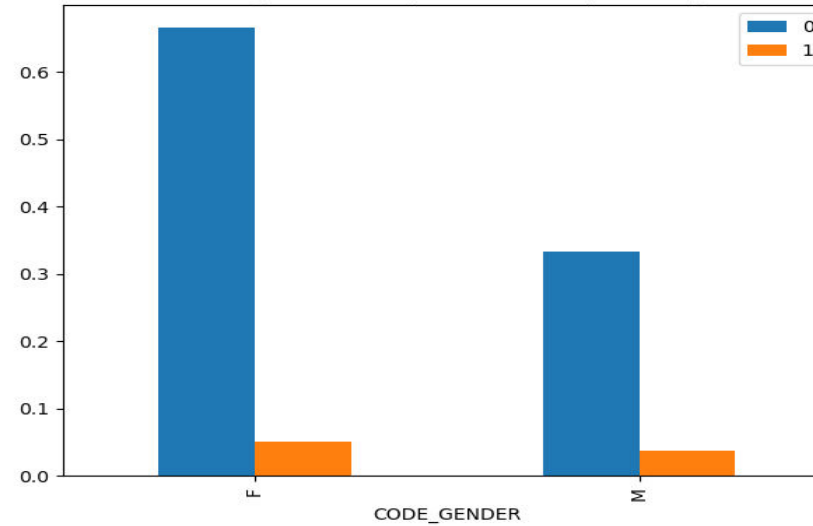
- NAME_CONTRACT_STATUS-We can observe that more than 60 % of previous applications are approved and Cancelled are just 18.9 %,where as refused 17.4% and unused contributes least percentage of 1.5.
- NAME_CONTRACT_TYPE-44 and 43 % of loans are Cash loans and Consumer loans.
- NAME_CLIENT_TYPE-More than 70% of clients are repeaters who have applied loans earlier.

ANALYSIS ON MERGED DATA (CODE_GENDER & NAME_CONTRACT_STATUS)

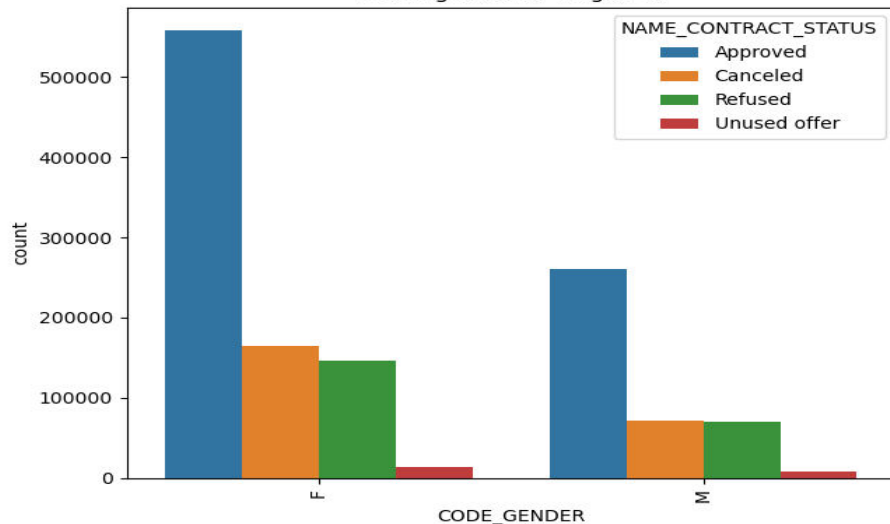
Plotting data for the column: CODE_GENDER



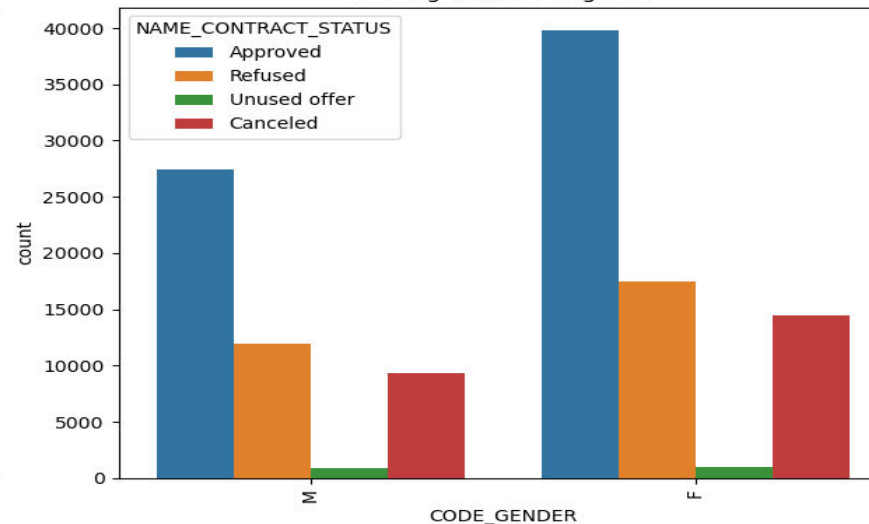
Plotting data for target in terms of percentage



Plotting data for Target=0



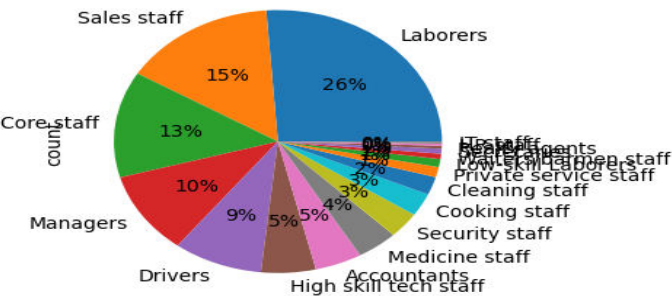
Plotting data for Target=1



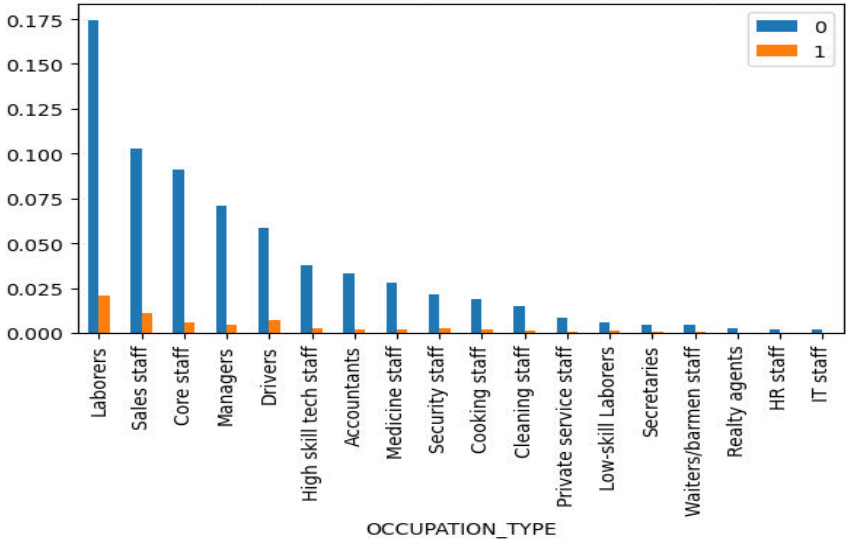
Approved Loans are approximately 40000 for Female and it is more than 25000 for Male in Target 1.

ANALYSIS ON MERGED DATA (OCCUPATION_TYPE & NAME_CONTRACT_STATUS)

Plotting data for the column: OCCUPATION_TYPE

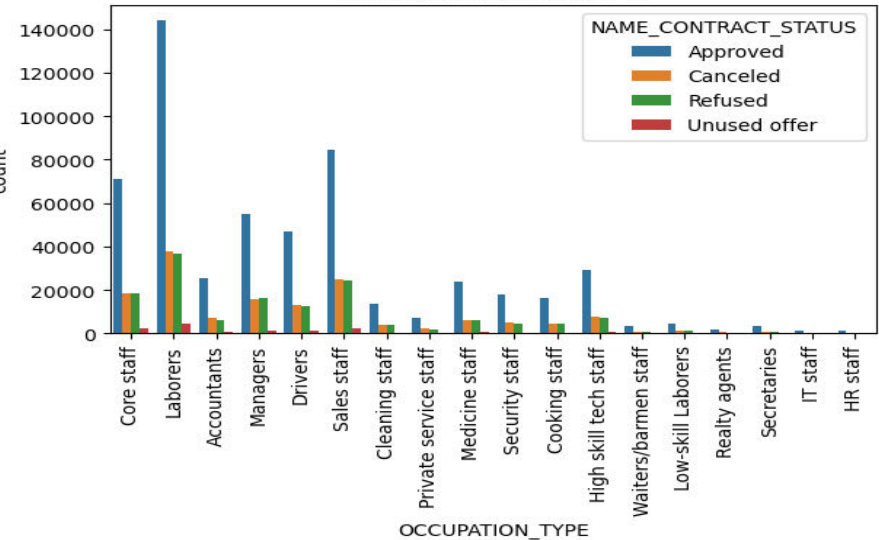


Plotting data for target in terms of percentage

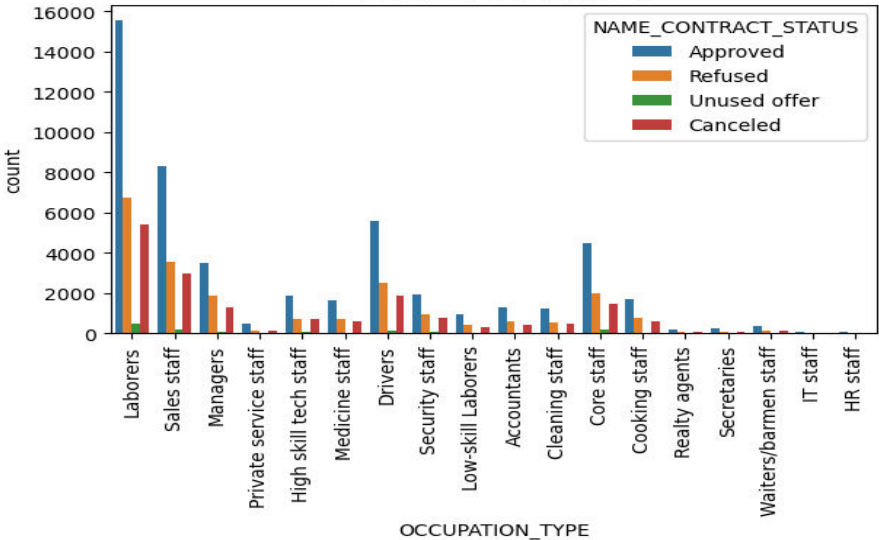


Laborers and Sales staff are having maximum approved loans in default category.

Plotting data for Target=0

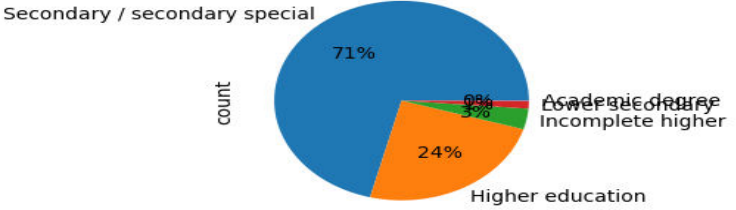


Plotting data for Target=1

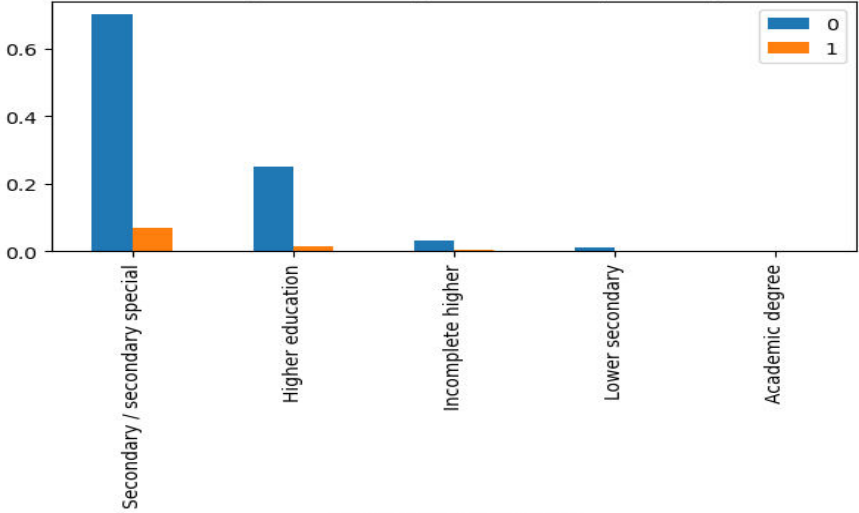


ANALYSIS ON MERGED DATA (AGE_GRP & NAME_CONTRACT_STATUS)

Plotting data for the column: NAME_EDUCATION_TYPE

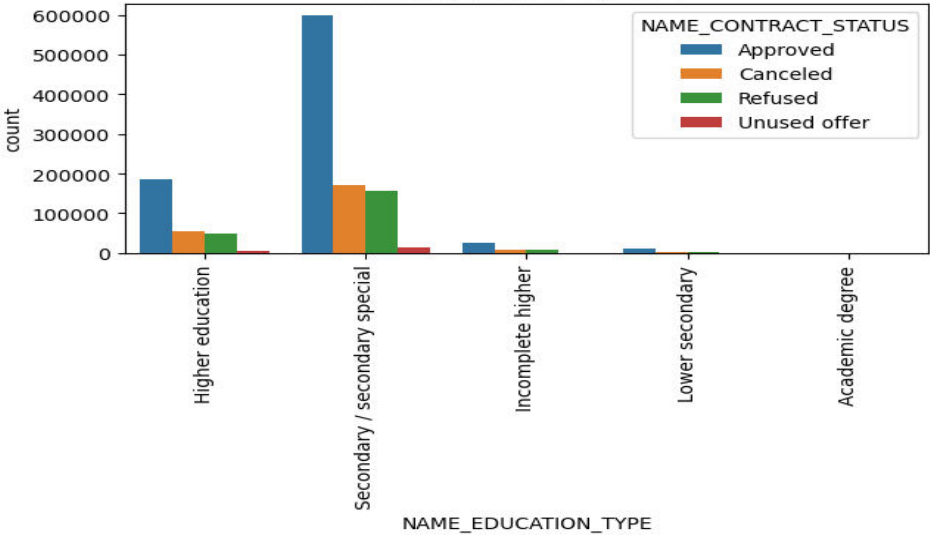


Plotting data for target in terms of percentage

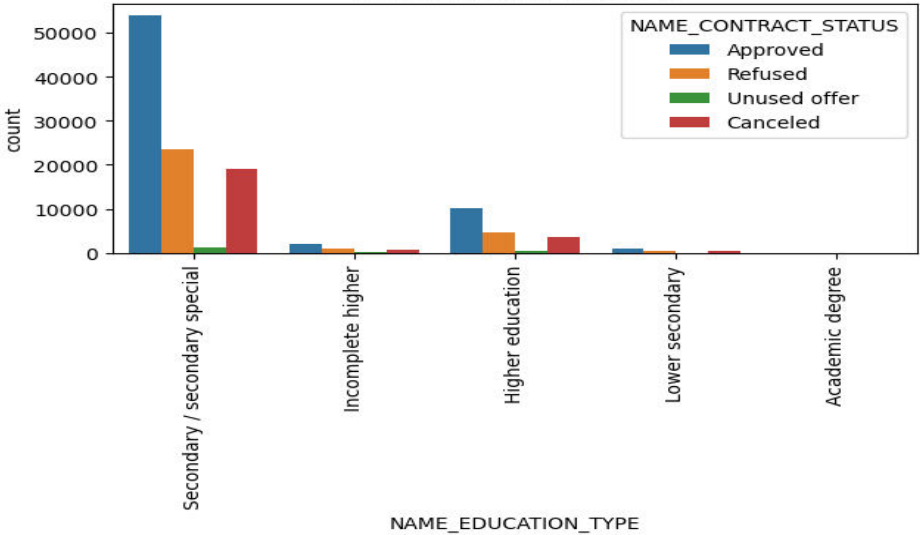


Secondary educated clients seems to be have their loans approved mostly in Target 1.

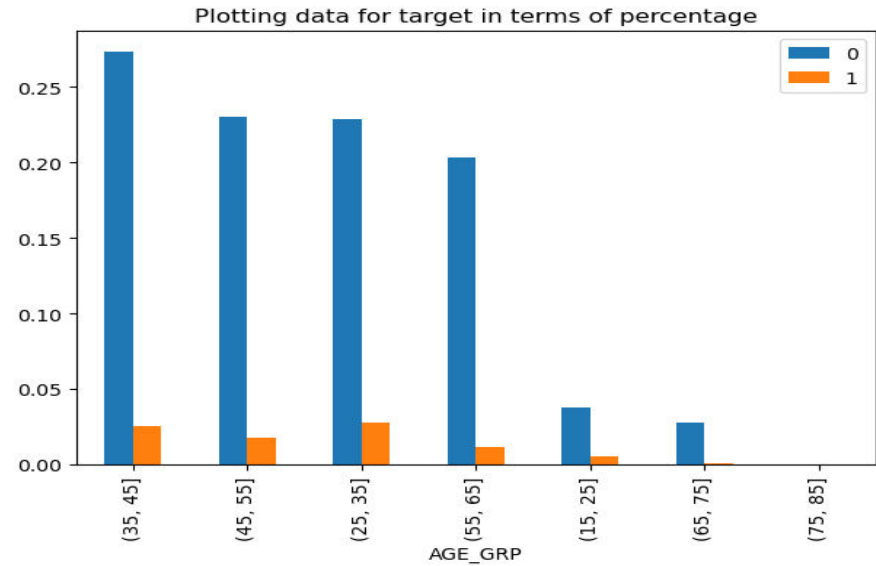
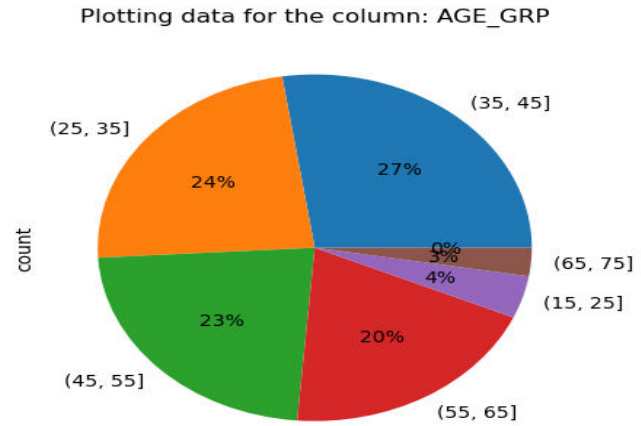
Plotting data for Target=0



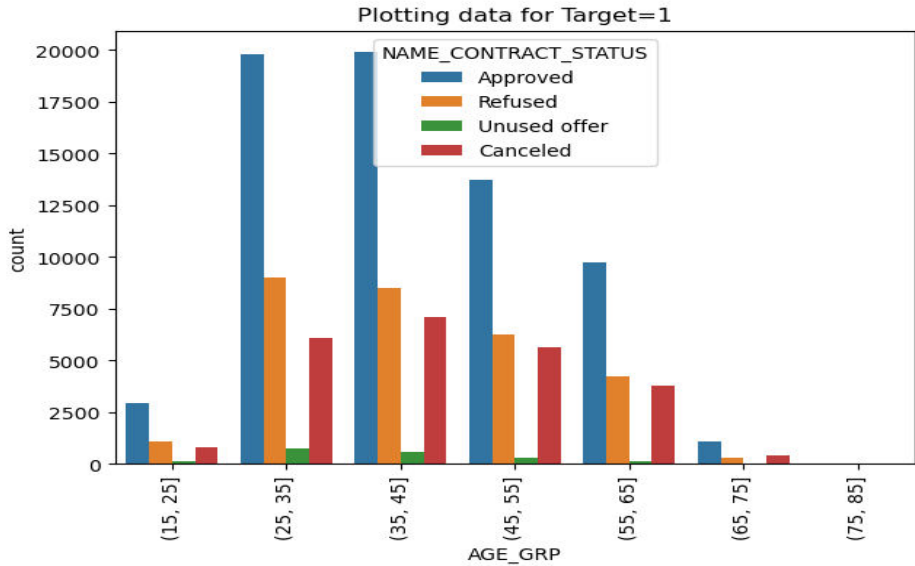
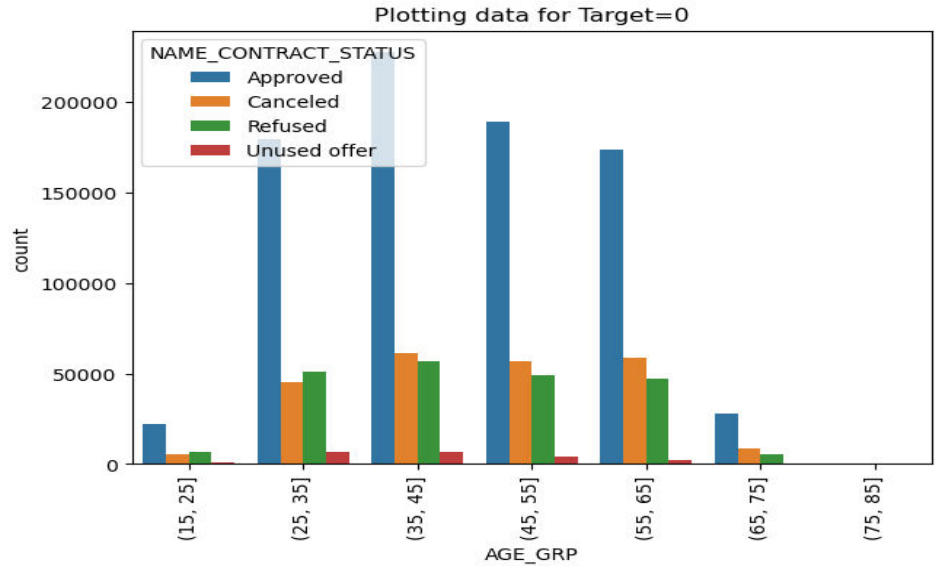
Plotting data for Target=1



ANALYSIS ON MERGED DATA (AGE_GRP & NAME_CONTRACT_STATUS)



Age grp 25-35 and 35-45 seems to be getting their loans approved and their number in defaulting is also high.





ANALYSIS CONCLUSIONS

- Male clients seem to be having more difficulty in repaying loans compared to female.
- Age group 25-35 and 35-45 are defaulting more compared to other age groups.
- Laborers and Sales staff are the occupation type to avoid while providing Loans.
- Most of the Loan applicants are repeaters who have applied loans earlier.
- Medium salary group are the income group to be taken care while approving loans as there are more number of defaulters.

THANK YOU

- SINIYYA YOUSUF